How-To-Do-It

The Scientific Method

An Instructor's Flow Chart

Ray T. Sterner

Recently, a number of authors have expressed concerns over scientific illiteracy in the United States (Culliton 1989; Ehrlich 1992; Gibbs & Lawson 1992; Levine 1990). Although 86% of high school students enroll in biology, only 12% take physics (U.S. Department of Education 1994). As an instructor, I have grappled with how to present and foster undergraduate interest in the principles/methods underlying scientific inquiry. This is not an easy task. Many students consider science boring and hard to understand; in the current vernacular, we're "pedagogically challenged"—challenged to stimulate and inform turned-off students about the scientific process.

A classroom technique that is designed to aid student understanding of science that I use involves a flow chart of the scientific method. The flow chart provides an excellent teaching aid for undergraduates; it affords a conceptual framework for organizing many research concepts taught in introductory science classes. Briefly, I hand out a schematic of the scientific method, then have students devise specific examples of each step during either an in-class or homework exercise.

The Scientific Method

Why are biology, chemistry, geology, physics, sociology, etc. considered sciences? Because these disciplines adhere to the scientific method (Christensen 1988). Although this is a rhetorical question that often baffles students, this definition is succinct, sufficient, and thought provoking. In short, the method specifies the question-and-answer process of science and outlines the sequence of steps needed to answer a question about a phenomenon—a guide to the scientist's decisions in conducting research. By adhering to the conventions of objective observation, control/balance of extraneous variance, and unbiased computation/analysis of data, the method ensures reproducibility of findings via empiricism.

Ray Sterner is a Research Psychologist (Animal Behavior) at the National Wildlife Research Center, USDA/APHIS, Ft. Collins, CO 80524-2719.

A Detailed Functional Flow Chart

Although Jessop (1970) and Lehner (1979) have published simplified schematics of the scientific method, detailed descriptions of the process are lacking. Here, I present a 17-step flow chart that standardizes the sequence of functional steps comprising the method (Figure 1). Of course, the conduct of research allows flexibility; adherence to all steps is not mandatory (Gibbs & Lawson 1992). My point is that scientists make decisions as they proceed. They choose the research approach, decide whether or not a pilot study/probe is needed, elect to replicate the study or not, and bring their own unique reasoning skills, ways of demonstrating effects, etc. to bear on the problem.

Researchers often combine/ignore certain steps or use multiple approaches simultaneously to address questions more quickly or efficiently. Thus, my flow chart is somewhat paradoxical; it standardizes a very unstandardized process but affords students insight into the way science operates. By charting the sequential steps ascribed to the scientific method, students can more easily grasp how theories/casual observations of phenomena lead to hypotheses, data gathering, and eventual support/refutation for explanations of events in the world around them.

Step 1: Ponder Theory or Observe a Phenomenon

The chart begins with the identification of a topic to be studied. A scientist may delve into the postulates of a theory or observe a thought-provoking phenomenon to conceive that a question exists. Theories are tentative explanations of how things occur and how phenomena may operate (e.g. Big Bang Theory, Theory of Evolution). These differ from models, which are analogies of known systems describing possible ways that other things may function [e.g. lock-and-key binding of neurotransmitters at post-synaptic sites, thermostatic (set-point) regulation of basal metabolism]. Although questions need not be derived from a theory, theories often guide researchers in predicting events or outcomes of research which ultimately support or deny the explanation.

Step 2: Ask Questions

Questions about observable events in the universe are the basis of the method. These encompass practically any phenomena that can be observed repeatedly (empiricism). Of course, religious or supernatural phenomena are precluded because they lack replication—a source of frequent contention between fundamentalist religious teachings and scientific inquiries (Erhlich 1992; Gillis 1994).

Step 3: Formulate Hypotheses (H_0 and H_1)

Borrowing from statistics, two types of hypotheses are invoked simultaneously: null (H_0) and alternative (H_1). H_0 and H_1 state that events will not change (not differ) and will change (differ), respectively, from some baseline/standard/control condition(s) due to the occurrence of an independent or naturally-occurring factor(s). Granted, students find this activity obtuse; the method involves indirect tests of hypotheses. H_0 is never accepted, and H_1 is never "proved." This reliance on rejection of H_0 as the way to answer questions about events with incomplete sampling is probably a major

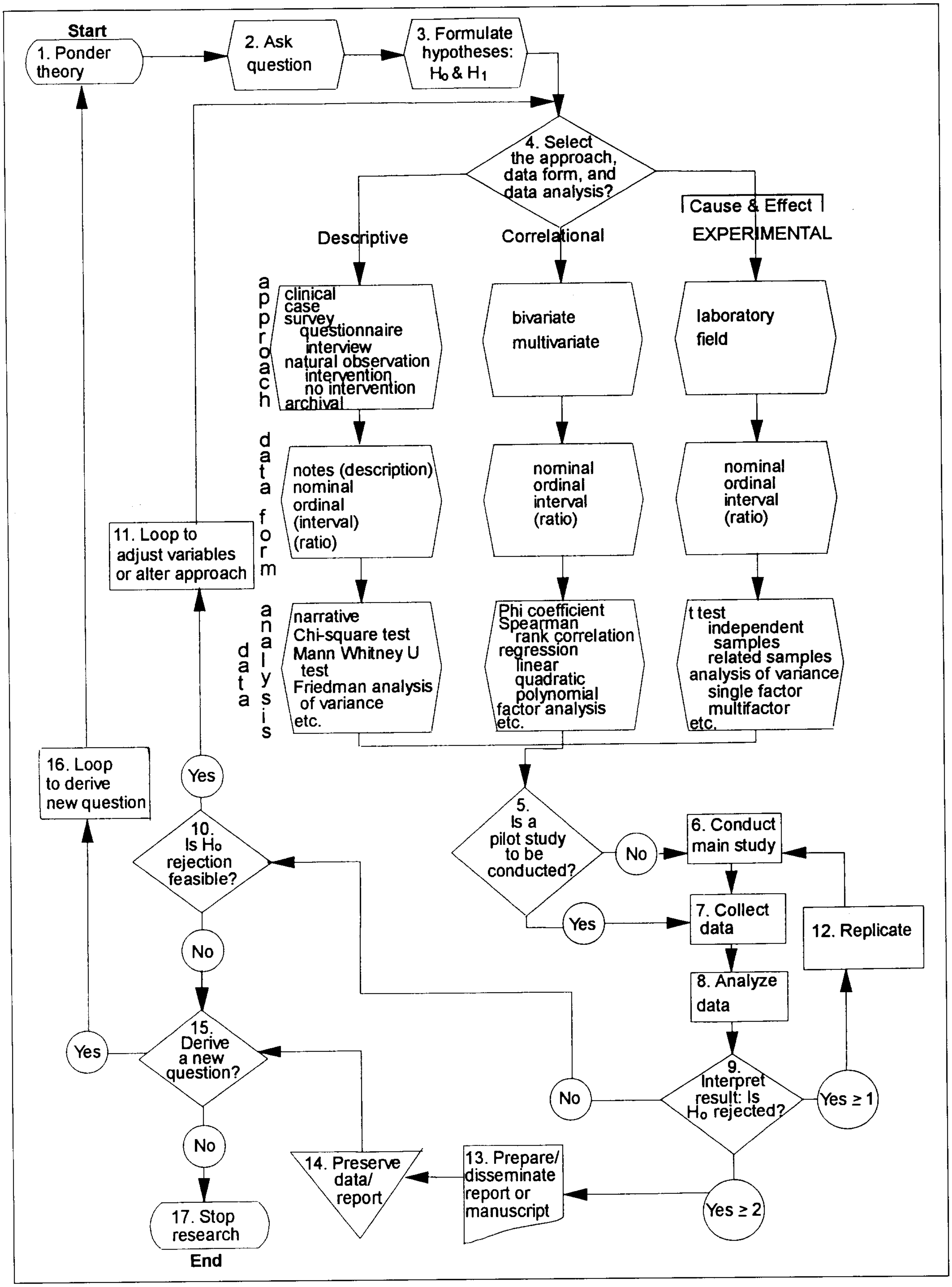

Figure 1. A 17-step, functional flow chart of the scientific method. (*Note*—Symbols are: ellipse = start/stop terminal, elongated hexagon = preparation step, rhombus = decision step, rectangle = predefined process, curved-base rectangle = document, inverted triangle = permanent storage, and circle = counter.)

source of undergraduate confusion about science.

A common definition of "hypothesis" given by Colorado undergraduates is "educated guess," after Kaskel, Hummer and Daniel (1985, p. 15). Although straightforward and simple, this definition ignores the mental labor required by scientists to formulate well-derived, explicit hypotheses. Moreover, it seems to trivialize the research process.

Step 4: Select the Approach, Data Form & Data Analysis

Descriptive, correlational, experimental, or a combination of these approaches are recognized ways of conducting scientific inquiry. These differ mainly in the control exerted over observations and in the determination of cause-and-effect. Only the experimental approach can derive cause-and-effect; the lack of random sampling and presence of uncontrolled sources of extraneous variance preclude cause-and-effect determinations using descriptive and correlational approaches.

a. Approaches *Descriptive research*—Clinical, case, survey (questionnaire and interview), natural observation (with and without intervention), and archival studies comprise possible descriptive accounts. These involve limited control over extraneous variance (i.e. fluctuations in measurements due to uncontrolled conditions/factors) and lack cause-and-effect inferences.

Most students are familiar with clinical-pharmaceutical trials (drug evaluations), medical-case studies, and psychological or marketing questionnaires and interviews. These involve tallies or descriptions of specific symptoms, etiologies, and responses of individuals.

Regarding natural observation, Jane van Lawick-Goodall's (1967) famous descriptions of chimpanzees (*Pan troglodytes*) and their social order is a classic example. Her observations of the chimps' reactions to a snake probably originated the "natural observation with intervention" sub-category; Van Lawick-Goodall (1967) reported the chimps' reactions to a snake as a serendipitous event that became a noteworthy observation. Scientists later introduced certain stimuli as a technique to structure natural observations.

Finally, archival studies involve searches of historical records. Paul Beier's (1991) report reviewing "53 known" incidents of cougar (*Felis concolor*) attacks upon humans in North America is great archival reading.

Correlational research—This approach assesses linear or curvalinear relationships between or among variables. Two categories of correlational studies are identified: bivariate and multivariate. Bivariate refers to the study of a linear or curvalinear relationship between two variables; multivariate refers to the assessment of linear or curvalinear associations among more than two variables at a time. Of course, relationships may be due to factors not directly measured in the study—correlation is not causation.

Experimental research—This type of research is characterized by "comparisons" among a set of experimental and baseline/control observations. Experimental groups/plots receive a manipulation by the scientist; control groups/plots receive exactly these same conditions, but lack the experimental manipulation or involve some sham-type manipulation—"a point of comparison." Herein is the logic for cause-and-effect; only one thing likely to affect the outcomes is changed at a time and the result(s) can be attributed to the manipulation(s). The design of experiments to effectively balance or control extraneous sources of variance is virtually a field of specialization unto itself.

Two types of experiments are recognized: laboratory and field. *Laboratory experiments* refer to manipulative studies conducted within a laboratory under precisely regulated conditions. *Field experiments* refer to manipulative studies carried out at specific geographical locations under existent environmental conditions.

b. Data Form Most textbooks recognize four types of measurement: nominal, ordinal, interval and ratio (Siegel 1956; Roscoe 1969). Selection of the data form involves tradeoffs—quantification of outcomes will be limited by the type of measurement used.

Nominal measurements—Nominal refers to assignments of observations to categories (male/female, slow/medium/fast); these allow judgments of equivalence—whether or not frequencies, percentages, etc. are roughly the same/different (equivalence/non-equivalence). Descriptive notes can also be a form of nominal measurement; described events can be categorized (e.g. juvenile/adult rodents by rapid/slow foraging).

Ordinal measurements—These measurements refer to ranked data; they allow judgments of equivalence and magnitude (equal to, greater than, or less than). Statistical tests can be used to draw inferences about the reliability of ordinal effects, but students must remember to state conclusions in greater than or less than terms, not score units or averages.

Interval measurements—Interval data involve the use of cardinal numbers and arithmetic operations (addition, subtraction, multiplication and division); equal units occur throughout the scores. These provide for judgments of equivalence and magnitude, plus arithmetic quantification of effects. Interval scales refer to equal units of scale with a relative zero point (e.g. water freezes at 0° C).

Ratio measurements—Ratio measurements have essentially the same properties as interval scales, but involve an absolute zero (e.g. absence of molecular motion at −273° C). Outcomes of data analyses may now be phrased in terms of multiplicative relationships and constants.

c. Data Analysis Detailed presentation of the myriad of data analysis considerations affecting the scientific method is beyond the scope of this paper (see SAS Institute 1987; Sokol & Rohlf 1969); moreover, this is perhaps the single-most confusing aspect of science for undergraduates. Data can be characterized using text or analyzed statistically—descriptively (e.g. mean, median, standard deviation), correlationally (e.g. Rank Order Correlation Coefficient, Factor Analysis), or inferentially (e.g. t-value, F-ratio). Ultimately, the analysis is determined by the approach, type of measurements, plus assumptions required of the statistical tests. Although prediction may be unaffected, tradeoffs of quantification and causation are sacrificed as the scientist employs more descriptive and correlational techniques.

Step 5: Is a Pilot Study To Be Conducted?

A pilot study allows the scientist to try out an idea without expending much effort. Some researchers conduct one or more pilot studies before performing a main study, while others don't bother with them. In general, a pilot study will benefit most research, and omission of these typically occurs when researchers have extensive information about the factors affecting a result; perhaps earlier studies have been reported/conducted, and this information obviates the need to "shake down" procedures or gain information about the utility of chosen procedures.

Step 6: Conduct Main Study

Step 6 is a predefined process. This means that the hypotheses, approach, data form, and data analysis (design) have been chosen. The main study is more extensive than a pilot test, involving larger, more thorough sets of observations.

Step 7: Collect Data

This step is also a predefined process; it involves the gathering of limited-pilot-study or extensive-main-study data. Avoidance of bias (e.g. neglecting to record certain events; counting experimental events differently than control events) is crucial here. Students must appreciate that objectivity drives the data collection. Without complete objectivity, biased/false results could mislead the scientist and other researchers in the years to come.

Today, laboratory and field data may be collected remotely using computers or data loggers, then deposited automatically into computer memory files. Although these procedures simplify data collection, the scientist retains responsibility for the collection of unbiased, accurate data.

Step 8: Analyze Data

Step 8 refers to conduct of the preselected analysis. As mentioned, this may consist of simple description or sophisticated multivariate statistical analysis of a set of observations, but the point remains that the scientist has previously determined how the data will be analyzed. Now, the analysis is simply performed on the collected data.

Step 9: Interpret Result: Is H_0 Rejected?

Interpretation of a result is complex, involving three possible levels of evaluation: data, statistics and results.

Data interpretation refers to an assessment of the raw data; this is the first level of analysis and requires internal checks of the data. The researcher must check for obvious errors of recording and transcripting. It is surprising how many studies are published with obvious errors in measurements and arithmetic. The occurrence of "floor" and "ceiling" effects (i.e. measurements near minimum and maximum values, respectively, that lack sensitivity) are also well known data analysis issues.

Statistical interpretation offers an intermediate analysis; this is tied closely to the assumptions and properties of the statistical test. Computation of a confidence interval to convey the degree of certainty that can be placed on the statistic (the probability that future samples will yield a similar statistic) is helpful. Typically, scientists use either 0.05 or 0.01 probabilities for statistics. This means that a particular statistic would occur less than five or one time(s) out of a hundred by chance, respectively.

Results interpretation is the most molar analysis. Despite the data and statistical analyses, the researcher must still decide about the real-world meaning of the data/statistics. Data/statistical significance need not equal biological, medical, chemical, economic, etc. significance.

In short, a decision concerning rejection or nonrejection of H_0, and conversely the non-acceptance or acceptance of H_1, is required here. If H_0 is rejected, a counter is set, and a decision to replicate or not occurs. A count of 1 nonrejection of H_0 leads to a second decision about the feasibility of H_0 and H_1 (Step 10: Is H_0 Rejection Feasible?), and, in turn, a decision about conducting a new line of investigation (Step 15: Derive a New Question); whereas, a count of 1 rejection of H_0 leads to replication (Step 12: Replicate). A count of ≥ 2 H_0 rejections is considered confirmation of a result; this leads to Step 13 (Prepare/Disseminate Report).

Step 10: Is H_0 Rejection Feasible?

Step 10 is controversial. Essentially, I contend that a scientist makes a decision whether or not s/he wants to pursue a particular line of research. That is, a judgment of the feasibility for attaining a "significant result" (reject H_0) by adjusting the independent variables or changing the approach exists. If yes, s/he goes to Step 11 and loops to adjust the variables or alter the research approach. If no, s/he goes to Step 15 and decides about the derivation of a new question or the stoppage of research (Step 17).

Step 11: Loop To Adjust Variables or Alter Approach

Entry into this loop allows the researcher to adjust or calibrate the variables before abandoning the hypotheses (e.g. increase or decrease the magnitude of the manipulation, select a new approach). In short, scientific results often depend upon proper calibration of the manipulated (independent) variable. For example, if a 10 mg/kg dose of an amphetamine drug yields no change in the activity of an organism, possibly 15 mg/kg should be used. A retest after adjustment allows a determination of whether the potential exists for the manipulation to influence the outcome (dependent) variable. Step 11 reflects a search by the scientist for the approach, factor, or intensity of variable that yields, correlates with, or causes a result.

Step 12: Replicate Study

Repeating Steps 5 (or 6) through 9 (inclusive) allows verification of a result within or between researchers, laboratories or environments. This would usually occur only if significant results were obtained after conduct of the main study; however, replication is a prerogative of the scientist. Replications can be conducted anytime using the same or new samples of elements, but generally new samples are used, particularly if the researcher is concerned that carryover (residual) effects could influence the outcome.

The "counter" displayed on this loop (Yes ≥ 1) indicates that at least one re-do of a main study is warranted for every set of significant results. Nevertheless, austere research budgets, coupled with professional pressures for scientists to publish, have led to avoidance of this step by many researchers. This also relates to popular criticisms levied against the sciences—many

CROSS-EYED CRICKET

studies reported in the news media don't withstand later replication. Of course, cost:benefit issues are important here, the reliability, validity, and costs of repeating studies must be weighed. Only via replication can scientists learn whether decisions relevant to the selection of variables, samples of objects, control of extraneous or nuisance variables, etc. will occur reliably.

Step 13: Prepare & Disseminate Report or Manuscript

Reports differ greatly in quality. Those submitted to scientific journals are usually "refereed" (critiqued) by other scientists in the field; however, a number of "non-referred journals and symposia" publish proceedings of meetings without peer review. Interestingly, many book publishers do not force mandatory compliance of content edits by peers—only grammatical/style edits are required for books. Although most scientific journals are reluctant to publish findings of no effect or negative results (i.e. the reason for inserting Steps 13 and 14 after successful replication of the study), some form of written, technical report should be prepared describing and interpreting procedures and findings from practically every study.

Step 14: Preserve Data/Report

Step 14 is an extension of Step 13 that involves the external, permanent storage of scientific information. Data and reports from all prior pilot and main studies should be preserved in permanent archives. As you might expect, the quality of archives varies widely; quality of storage facilities and the diligence expended on assembly of materials differ among scientists, archivists and institutions. Archive files of completed studies (i.e. protocols, reports, raw data, correspondence, etc.) are maintained by many government-sponsored research laboratories, with most academic institutions relying solely upon each faculty member to maintain such files.

Recent Good Laboratory Practice (GLP) and Good Clinical Practice (GCP) regulations for federally-registered chemicals (U.S. Environmental Protection Agency 1994) and drugs (U.S. Food and Drug Administration 1994) require quality assurance (QA) inspections of research procedures and archives of data for the life of the chemicals'/drugs' registrations. Use of acid-free paper for data/reports and storage in locked, fireproof cabinets/rooms are key concerns of archivists.

Step 15: Derive a New Question?

This is the final decision point in the chart. It affords continuance of the investigations (Step 16: Loop to Derive New Question) or cessation of the process (Step 17: Stop Research).

Step 16: Loop To Derive or Test a New Question

This step refers to a loop for deriving/testing a new question—the sequence of steps in the method is reinitiated (Steps 1–15 are performed again). The scientist may pursue other hypotheses related to the theory or possibly begin a new line of research (identify critical test points in some other theory). The search for improved understanding continues.

Step 17: Stop Research

Finally, Step 17 avoids creation of an endless loop. It recognizes that the scientist either has a choice to cease research (e.g. take a temporary hiatus from research or permanently retire) or s/he may become incapacitated at any time (e.g. become ill, die).

Conclusions

The scientific method refers to the general technique used by many researchers to gather information about phenomena in the universe. It is the stepwise sequence whereby researchers ponder questions, formulate hypotheses, submit these hypotheses to empirical test (i.e. select approaches, design/conduct studies, collect data, analyze data, draw inferences, and document/store findings) to gain objective information (answers). "The method" literally "makes a science a science." Although descriptive and correlational studies afford new insights concerning key factors that affect observed events, only experiments provide cause-and-effect determinations. Through test and replication, adherence to the scientific method implies that ultimately only consistent, valid information will be accepted into the body of scientific knowledge.

Acknowledgments

I thank Philip Lehner, Donald Elias, Russell Mason, and Kathleen Fagerstone for helpful reviews of the manuscript; Dolores Steffen prepared Figure 1.

References

Beier, P. (1991). Cougar attacks on humans in the United States and Canada. *Wildlife Society Bulletin, 19,* 403–412.

Christensen, L.B. (1988). *Experimental Methodology.* Boston, MA: Allyn & Bacon.

Culliton, B.J. (1989). The dismal state of scientific literacy. *Science, 243,* 600.

Erhlich, P.R. (1992). One ecologist's opinion on the so-called Stanford scandals and social responsibility. *Bioscience, 42,* 702–705.

Gibbs, A. & Lawson, A.E. (1992). The nature of scientific thinking as reflected by the work of biologists & by biology textbooks. *The American Biology Teacher, 54*(3), 137–152.

Gillis, A.M. (1994). Keeping creationism out of the classroom. *Bioscience, 44,* 650–660.

Jessop, N.M. (1970). *Biosphere: A Study of Life.* Englewood Cliffs, NJ: Prentice-Hall.

Kaskel, A., Hummer, P.J. & Daniel, L. (1985). *Biology: An Everyday Experience. Teacher's Annotated Edition.* New York: Charles Merrill.

Lehner, P.N. (1979). *Handbook of Ethological Methods.* New York: Garland STPM Press. pp. 9–16.

Levine, J. (1990). Scientific illiteracy: We have met the enemy and he is us. *Oceanus, 33,* 66–73.

Roscoe, J.T. (1969). *Fundamental Research Statistics for the Behavioral Sciences.* New York: Holt, Rinehart and Winston. pp. 86–87.

Siegel, S. (1956). *Nonparametric Statistics for the Behavioral Sciences.* New York: McGraw-Hill. pp. 6–34.

SAS Institute, Inc. (1987). *SAS/STAT" Guide for Personal Computers.* Cary, NC: SAS Institute. 1029 pp.

Sokol, R.R. & Rohlf, F.J. (1969). *Biometry: The Principles and Practices of Statistics in Biological Research.* San Francisco: Freeman & Co.

U.S. Department of Education. (1994). Digest of education statistics. Page 135 in *Office of Educational Research and Improvement, National Center for Education Statistics (NCES 94–115),* Washington, DC: U.S. Government Printing Office.

U.S. Environmental Protection Agency. (1994). *Code of Federal Regulations: Protection of Environment. 40 (parts 160–189).* Washington, DC: U.S. Government Printing Office. 680 pp.

U.S. Food and Drug Administration. (1994). *Code of Federal Regulations: Food and Drugs. 21 (parts 1–99).* Washington, DC: U.S. Government Printing Office. 420 pp.

van Lawick-Goodall, J. (1967). *My Friends the Wild Chimpanzees.* Washington, DC: National Geographic Society. 204 pp.